Running Head: EFFECTS OF ITEM WORDING

Detecting Effects of Positively and Negatively Worded Items on a Self-Concept Scale for third and sixth grade Elementary Students

Yi-Hsin Chen, Gianna Rendina-Gobioff, and Robert F. Dedrick

University of South Florida

Paper presented at the annual meeting of the Florida Educational Research Association,

Tampa, FL, November 14-16, 2007

Please address correspondence concerning this manuscript to:

Yi-Hsin Chen

Department of Educational Measurement and Research

University of South Florida

4202 E. Fowler Ave. EDU 162

Tampa, FL 33620-7750

Email: ychen@coedu.usf.edu

Abstract

Method effects associated with item wording have been explored in a variety of instruments and found that the practice of using positively- and negatively- worded items may introduce systematic measurement errors that disrupt analyses and interpretations of the results. Therefore, the first purpose in the present study was to explore if method effects were present in a Chinese general self-concept scale, originally developed in Chinese. The second purpose was to determine if the factor structure of the method effects, if present, differed for third and sixth grade students. The third purpose was to determine if the observed method effects were related to other substantively meaningful variables.

Results from a series of CFAs support the presence of method effects associated with the negatively and positively worded items and method effects were largest for the negatively worded items. The results from multigroup model comparisons indicate that the factorial structure of these method effects was not significantly different for third and sixth graders. Three demographic variables, including student gender, student grade level, and students' overall performance ratings provided by teachers, were used to examine the relationships with negative method effects. The results of path analysis indicated that students who were rated lower by their teachers were more likely to endorse negative statements about themselves. Furthermore, students in grade 3 were significantly more likely to endorse negative statements compared to students in grade 6. But gender was not significantly related to the negative method factor.

Detecting Effects of Positively and Negatively Worded Items in a Self-Concept Scale among Elementary Students

The use of both positively- and negatively- worded items in survey instruments has been advocated for many decades (e.g., Anastasi, 1982; Anderson, 1981; Mehrens & Lehmann, 1983; Nunnally, 1978; Spector, 1992) to avoid response bias such as acquiescence or agreement bias (DiStefano & Motl, 2006; Cronbach, 1950; DeVellis, 1991), which is the tendency to agree with the survey items, independent of item content. Negatively worded items are included to act as "cognitive speed bumps that require respondents to engage in more controlled, as opposed to automatic, cognitive processing" (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003, p. 884). The recommendation to word approximately half of the items positively and half negatively so as to minimize response bias is based on two crucial and implicit assumptions. First, the items worded in the opposite ways are measuring the same construct. Second, respondents are able to reply to positively and negatively worded items equivalently (Benson & Hocevar, 1985; Marsh, 1996).

Many empirical studies involving psychometric analyses of survey instruments suggest that the practice of using positively- and negatively- worded items may introduce systematic measurement errors that disrupt analyses and interpretations of the results (Horan, Distefano & Motl, 2003; DiStefano & Motl, 2006; Quilty, Oakman, & Risko, 2006). Constructs that are conceptualized as unidimensional may appear to be multidimensional when positively and negatively worded items are included (i.e., positively and negatively worded items form separate factors). The impact of negatively worded items on the measurement of participants' responses represents one type of method effect (see Podsakoff, MacKenzie, Lee, & Podsakoff, 2003 for a review of common method effects). Method effects represent the systematic variance introduced as a result of the research approach or method used to measure the trait under investigation.

Campbell and Fiske (1959) called attention to method effects almost 50 years ago, arguing that "each test or task employed for measurement purposes is a trait-method unit, a union of a particular trait content with measurement procedures not specific to that content. The systematic variance among test scores can be due to responses to the measurement features as well as responses to the trait content" (p. 81).

Method effects associated with item wording have been explored in a variety of instruments including those that measure self-esteem, loneliness, affect, and organizational commitment (e.g., Bolin & Dodder, 1990; Kelloway, Catano, & Southwell, 1992; Marsh, 1996; Roberts, Lewinsohn, & Seeley, 1993). The Rosenberg Self-Esteem (RSE) scale (Rosenberg, 1965), one of the most widely used self-esteem measures, has been the focus of several studies that have examined method effects. Originally, there were 10 items on this scale, including 5 positively and 5 negatively phrased items. This scale was created to assess a general self-esteem factor. In the early investigations of the RSE scale, exploratory factor analysis was applied to examine its factor structure (e.g., Carmines & Zeller, 1979; Hensley & Roberts, 1976). Two factors reflecting positively and negatively worded items were found in these studies, but they were not substantively meaningful. The authors interpreted the two factors as method effects resulting from item phrasing. To date, more and more researchers have suggested applying confirmatory factor analysis (CFA) to examine item wording effects (DiStefano & Motl, 2006; Horan, DiStefano, & Motl, 2003; Marsh, 1986; 1996; Marsh & Grayson, 1995; Tomás & Oliver, 1999) so as to overcome the shortcomings of exploratory factor analysis, which is unable to differentiate competing factor structures of the scale (Marsh, 1996).

Research studies using CFA to evaluate method effects resulting from negatively and positively worded items in the Rosenberg Self-Esteem (RSE) scale have posited a number of

alternative measurement models underlying the RSE (Marsh, 1996; Motl & DiStefano, 2002; Tomás & Oliver, 1999). These models include: (a) a one-factor model representing general self-esteem, (b) a two-factor model representing positive and negative self-esteem as two substantive, correlated factors, (c) a two-factor model with one factor representing general self-esteem and one factor representing a negative method effect (general self-esteem and the negative method factor are assumed to be uncorrelated), (d) a two-factor model with one factor representing general self-esteem and one factor representing a positive method effect (general self-esteem and the positive method factor are assumed to be uncorrelated, (e) a one-factor model representing general self-esteem with a negative method effect modeled using correlated uniquenesses (error terms) for the negatively worded items, and (f) a one-factor model representing general self-esteem with a positive method effect modeled using correlated uniquenesses (error terms) for the positively worded items.

Results of confirmatory factor analyses from several studies that have examined these alternative models have consistently found statistically significant method effects associated with negatively worded items. These effects have been found with middle school (Marsh, 1996) and college students (DiStefano & Motl, 2006) and with instruments that have been translated and adapted from English to Spanish (Tomás & Oliver, 1999). Building on these previous research studies, the present study explored if method effects were present in a Chinese general self-concept scale used to collect Taiwanese elementary student responses. This scale was not translated from an English version of a self-concept survey, but rather was originally developed in Chinese for school students. The second purpose of this study was to determine if the factor structure of the method effects, if present, differed for third and sixth grade students. Previous research has suggested that because of developmental factors, younger students and students

with lower verbal abilities may have difficulties with negatively worded items (Marsh, 1986). Although this research has identified method effects in various age groups, multigroup CFA has not been used to evaluate the invariance of these method effects. A third purpose of the study was to determine if the observed method effects were related to other substantively meaningful variables. Quilty, Oakman, and Risko (2006), for example, found that college students who had higher levels of the personality trait of avoidance motivation were significantly more likely to endorse negatively worded items. In a follow up study Quilty, Oakman, and Risko (2006) found that individuals in an adult community sample who were more conscientious and emotionally stable were less likely to endorse negatively worded items. The current study examined whether male and female students responded differently to negatively worded items and if students who were rated highly by their teachers on overall school performance responded differently to negatively worded items.

Methods

Participants

Stratified sampling was used to sample students by classroom in 10 schools from four cities in Taiwan. Participants consisted of 752 elementary school students. There were 324 (43%) students in grade 3 and 428 (57%) students in grade 6. The sample was 51.9% male and 48.1% female.

Instrument

Self-concept. Eight items from a Chinese general self-concept scale (Table 1) for elementary students were extracted from the 30-item 3-factor school self-concept scale developed by Chang, Chen, Chen, and Jang (1997). Designed for Taiwan's elementary school context, the original school self-concept scale was created in Chinese. The Chinese general self-

concept scale used in the present study contained eight items with four positively worded items and four negatively worded. A balanced design consisting of an equal number of positively and negatively worded items was used to obtain a more valid index of method effects (Benson & Hocevar, 1985). Positive (e.g., I am happy) and negative (e.g., I am not happy) item formats were used in the scale in order to obtain reliable and accurate responses from students (Schriesheim, Eisenbach, & Hill, 1991). Students responded to the items on this questionnaire using a 4-point scale to rate the extent of their agreement (4=strongly agree; 3=agree; 2=disagree; 1=strongly disagree). Negatively worded items were reverse-scored. Cronbach alpha reliability coefficients for the eight-item scale for the total sample and for grades three and six were .61, .56, and .65, respectively. For the four positively worded items the Cronbach alphas for the total sample and for grades three and six were .67, .63, and .67, respectively. For the four negatively worded items the Cronbach alphas for the total sample and for grades three and six were .54, .47, and .61, respectively.

Insert Table 1

about here

Teachers' rating of overall performance. One variable that was examined as a potential correlate of method effects associated with the self-concept measure was teachers' ratings of students' overall performance in school. Teachers' ratings of students were based on their perceptions of a student's overall school performance, including both academic and nonacademic performances and behaviors. Academic performance encompassed test and project grades, homework efforts, and attitudes toward learning. Nonacademic performance

included behaviors such as bringing learning materials, relationships with classmates, and obeying school or class rules. Each homeroom teacher was asked to nominate from his/her class (maximum number of students in classes ranged from 20 to 45) at most seven students with the best overall performance, forming the high performance group (rated 1), and at most seven students with the worst overall performance (rated 3), forming the low performance group. Those students not named to either of these groups were in the average performance group (rated 2).

Statistical Analyses

There were four analytical procedures conducted in this study, including: (a) exploratory factor analysis (EFA) of the 8-item general self-concept scale, (b) a series of CFA models, (c) multigroup CFA analyses for the 3rd and 6th grade students, and (d) a path analysis that was used to examine the relation between students' gender, grade, and their overall performance rating provided by their teachers (predictor variables) and method effects (dependent variable).

Exploratory factor analysis was applied to examine how the general self-concept items loaded onto latent factor(s) (e.g., one general factor or two positive and negative factors). Principal axis factoring extraction and promax rotation were implemented using SPSS 15.0 statistical software.

The second phase used a series of CFAs that have been previously used in the literature (e.g., Distefano & Motl, 2006; Marsh 1996; Horan, DiStenfano, & Motl, 2003; Marsh & Grayson, 1995; Tomás & Oliver, 1999). Six CFA models were examined in this study (see Figures 1a – 1f). Model 1a was a one-factor model consisting of a general self-concept factor. Model 1b contained two oblique positive and negative self-concept factors. Models 1c through 1f were models that contained one general self-concept factor representing all items plus different

method effects. Model 1c added a negative method factor with all negatively worded items. Model 1d added a positive method factor with all positively worded items. Model 1e added correlated uniquenesses among all negatively worded items to model the negative method factor. Model 1f added correlated uniquenesses among all positive items to model the positive method factor. The CFAs were conducted using Mplus version 4.2 (Muthén & Muthén, 1998-2004). Maximum likelihood estimation was used. Models were evaluated using the following fit indices: the chi-square statistic, Comparative Fit Index (CFI), root mean squared error of approximation (RMSEA), and standardized root mean residual (SRMR). Hu and Bentler's (1999) cutoff values of greater than or equal to .95 for the CFI, less than or equal to .08 for the SRMR, and less than or equal to .06 for the RMSEA were used as general indicators of acceptable fit of the models; however, substantive issues such as the interpretability of the parameter estimates were also considered.

Insert Figures 1a – 1f

about here

The third phase of statistical analysis investigated whether method effects identified in the second phase would produce different parameter estimates across grade levels. Thus the best fitting CFA model from the second phase was compared for 3rd and 6th graders. Multigroup confirmatory factor analysis (CFA) was used to test the equality of parameter estimates across grade. Equality of parameter estimates was tested using a series of hierarchically ordered models of increasing restrictiveness. The strategy used to evaluate the various levels of measurement

invariance was to compare the nested likelihood ratio χ^2 difference ($\Delta \chi^2$) relative to the difference in the degrees of freedom (Δdf) for the models being compared. These tests were supplemented by comparing the changes in the CFI and SRMR along with their actual values to determine if the equality constraints produced unacceptable fit based on the guidelines by Hu and Bentler (1999).

In the fourth phase of the analysis, a path model was used to examine the relation between students' gender, grade level, and their overall performance rating provided by their teachers and the method effects.

Results

Descriptive Statistics

Table 1 provides descriptive statistics for the individual items and summary scores for the Chinese general Self-Concept Scale (all eight items, plus separate scores for the four positive and four negative items). The largest values for skewness occurred for negatively worded item 6 for the total sample (skewness = -1.06) and for positively stated items 1 and 2 for the third grade sample (skewness values = -1.15 and -1.10, respectively).

Exploratory Factor Analysis

The scree plot (Figure 2) from the exploratory factor analysis showed that there were only two components with eigenvalues greater than one. This output indicated that a two-factor solution was appropriate for the Chinese general self-concept scale. Pattern coefficients (i.e., standardized regression coefficients) in Table 2 further show that all positively worded items had high factor loadings on Factor 1 (named the Positive factor), but low factor loadings on Factor 2 (named the Negative factor). All negatively worded items had the opposite pattern, which had high factor loadings on the negative factor, but low factor loadings on the positive factor. Results

from exploratory factor analysis provided evidence that the self-report scale involving positively and negatively worded items yielded a two-factor solution, consisting of a Positive factor and a Negative factor.

Insert Figure 2 and Table 2

about here

Confirmatory Factor Analysis: Models 1a – 1f

The goodness of fit indices and the standardized parameter estimates for each of the six confirmatory factor analysis models are presented in Tables 3 and 4.

Model 1a. The fit of the one-factor self-concept model was not acceptable (CFI = .699, RMSEA = .116, SRMR = .090). Standardized factor loadings ranged from .10 (negative item 7) to .63 (positive item 4).

Model 1b. Model-data fit of the two-factor self-concept model containing a positive self-concept factor and a negative self-concept factor was good (CFI = .962, RMSEA = .048, SRMR = .037). Standardized factor loadings for the items within the positive factor ranged from .52 (positive item 2) to .65 (positive item 4), and for the items within the negative factor from .39 (negative item 5) to .60 (negative item 8). The correlation between the positive and negative factors was .26 ($p < .01$).

Model 1c. This two-factor model, consisting of one general self-concept factor and a method factor representing the negatively worded items. had good fit (CFI = .965, RMSEA = .044, SRMR = .028). Standardized factor loadings for the eight items on general self-concept

ranged from .04 (negative item 7) to .65 (positive item 4). Loadings for the items within the negative method factor ranged from .37 (negative item 5) to .61 (negative item 8). Note that the correlation between the general self-concept and negative method factors was fixed to zero.

Model 1d. This two-factor model, similar to Model 1c except that the method factor was represented by the positively worded items, did not fit as well as Model 1c but was good (CFI = .951, RMSEA = .052, SRMR = .035). Standardized factor loadings for the eight items on general self-concept ranged from .13 (positive item 2) to .58 (negative item 8). Loadings for the items within the positive method factor ranged from .50 (positive item 2) to .64 (positive item 4). Note that the correlation between the general self-concept and positive method factors was fixed to zero.

Model 1e. This model consisted of one factor representing general self-concept. The negative method effects were modeled by including correlated uniquenesses for the pairs of negatively worded items. Model-data fit of the model was good (CFI = .969, RMSEA = .044, SRMR = .026). Standardized factor loadings for the eight items on general self-concept ranged from .04 (negative item 7) to .65 (positive item 4). All six pairs of correlated uniquenesses (errors) were statistically significant ($p < .01$) and ranged from .15 (negative item 1 with negative item 3) to .29 (negative item 3 with negative item 4).

Model 1f. This model was identical to Model 1e except that the positive method effects were modeled by including correlated uniquenesses for the pairs of positively worded items. Model-data fit was good (CFI = .959, RMSEA = .051, SRMR = .033) but not as good as Model 1e. Standardized factor loadings for the eight items on general self-concept ranged from .13 (positive item 2) to .58 (negative item 8). All six pairs of correlated uniquenesses (errors) were statistically significant ($p < .01$) and ranged from .25 (positive item 1 with positive item 2) to .39

(positive item 1 with positive item 4).

Results of this series of CFAs support the presence of method effects associated with the negatively and positively worded items (see Tables 3 and 4). Model 1e (one general factor with correlated uniquenesses for the negatively worded items) was determined to have the best fit based on the CFI, RMSEA, and SRMR (Table 3). Consistent with previous research (DiStefano & Motl, 2006), method effects were largest for the negatively worded items.

Insert Tables 3 and 4

about here

Confirmatory Factor Analysis: Multigroup Models

Model 1e (one general factor with correlated uniquenesses for the negatively worded items) was used as the baseline model in testing the invariance of the method effects across third and sixth grade Taiwanese elementary students. This model was fit to each group's data without any equality constraints. The fit of the multigroup model was good (CFI = .985, RMSEA = .031, SRMR = .027). To evaluate the equality of the factor loadings (pattern coefficients) across groups, the change in the χ^2 ($\Delta \chi^2$) relative to the change in the degrees of freedom (Δdf) for the nested models was examined for Model 2 (equal factor pattern coefficients) compared to Model 1 (no equality constraints). The $\Delta \chi^2$ was 6.39 ($\Delta df = 7$), which was not statistically significant, suggesting that the assumption of equal loadings across groups was tenable. To test the equality of the correlated uniqueness Model 3 (equal covariances between uniquenesses) was compared to Model 2 (covariances between uniquenesses freely estimated). The $\Delta \chi^2$ was 4.78 ($\Delta df = 6$),

which was not statistically significant, suggesting that the assumption of equal correlated uniquenesses across groups was tenable. These results indicate that the factorial structures of these method effects were not significantly different for third and sixth graders (see Table 5 for a summary of the invariance tests).

Insert Table 5

about here

Confirmatory Factor Analysis:Path Analysis

Although Model 1e (one general factor with correlated uniquenesses for the negatively worded items) provided the best fit of the data and an estimate of method effects associated with the negatively worded items, in order to address the last purpose of the study (i.e., predictors of method effects), it was necessary to model the negative method effects as a separate latent variable. Model 1c, which consisted of a General Self-Concept factor and a Negative method factor had excellent fit and therefore was used for this purpose. Prior to using this model, multigroup CFA was used to evaluate the equivalence of the factor loadings for the general self-concept and negative method factors across grade levels. Table 6 summarizes the series of models. Overall, these results, which parallel those of Model 1e, indicated no statistically significant differences in the loadings across grades.

Insert Table 6

about here

Figure 3 displays the model that was used to examine the relation between three demographic variables, including student gender, student grade level, and students' overall performance ratings provided by teachers, and negative method effects. The standardized regression coefficient between students' overall performance rating and the negative methods factor was statistically significant (beta = -.265, $p < .01$) and indicated that students who were rated lower by their teachers were more likely to endorse negative statements about themselves. Students in grade 3 were significantly more likely to endorse negative statements compared to students in grade 6 (beta = .195, $p < .01$). Gender was not significantly related to the negative method factor (beta = .065, $p > .05$).

Discussion

Method effects may reduce the construct validity of the scores from survey instruments and distort the observed relations between measures, and therefore, it is important for researchers to understand the degree to which these effects are present. The present study identified method effects associated with the use of negatively worded items with a researcher-developed general self-concept measure that was developed in Chinese and used with a sample of children from Taiwan. Results of this study parallel those of a large cross-cultural study that investigated the use of the Rosenberg Self-Esteem Scale, translated into 28 different languages, and used with adults in 53 countries (Schmitt & Allik, 2005). In their study, Schmitt and Allik found that across many of the countries, individuals responded differently to the negatively and positively worded items.

Although previous studies have examined method effects for instruments used with different age groups, there has been only one other study, to our knowledge, that has examined the factorial invariance of method effects. Motl and DiStefano (2002), in their longitudinal

analysis of a global self-esteem measure that was used as part of the National Educational Longitudinal Study (NELS) found that over time (1988, 1990, and 1992) the factor structure of the method effects associated with the three negatively worded items (out of seven) was invariant. Because of the stability of these method effects, Motl and DiStefano suggested that method effects should "be considered of potential substantive importance rather than simply substantively irrelevant noise" (p. 571). In the present study, multigroup confirmatory factor analysis was used with cross-sectional data to examine method effects with third and sixth grade students. Similar to Motl and DiStefano (2002), the present study found that the models representing method effects associated with the negatively worded items (Model 1e: General factor + Correlated errors for negative items, and Model 1c: General factor + Negative Method factor for negative items) were not significantly different in their factorial structure for third and sixth graders. This consistency in the structure of the method effects provides additional support of the potential substantive meaningfulness of method effects.

With these findings, a question that arises is what factors may play a role in individuals responding differently to negatively worded items? In an attempt to address this question, the present study examined the relations between students' gender, grade level, and their overall performance rating provided by their teachers and the method effects associated with the negatively worded items. Results suggested that there were significant relations between students' grade levels and their ratings provided by teachers and students' tendency to respond to negatively worded items. Previous research by Marsh (1986) found that younger students had difficulties with negatively worded items. Future research, using one-on-one cognitive interviews, may provide answers for why students in the current study in grade 3 were significantly more likely to endorse negative statements compared to grade 6 students, and also

why students who were rated lower performance by their teachers were more likely to endorse negative statements about themselves.

Methodologically, the present study used a series of CFA models to evaluate competing structures underlying the self-concept scale. Model 1e (General factor + Correlated errors for negative items) and Model 1c (General factor + Negative Method factor for negative items) provided the best fit to the data and each identified the negatively worded items as a source of measurement error. One of the advantages of Model 1e (correlated uniqueness model) is that in the estimation process it rarely leads to inadmissible solutions. The disadvantage of this model is that the negative wording effects are not modeled as a distinct latent variable. Model 1c models the negatively worded items as a latent variable and as a result it is possible to examine the relationship of this variable with other theoretically meaningful variables.

Previous research has examined personality variables and other response styles (e.g., social desirability) and their relation to method effects represented by a distinct latent variable. In the present study, three demographic variables and their relation to method effects were examined. Future research should expand on this list of variables and include additional demographic variables (e.g., race/ethnicity) and contextual variables (mode of administration such as face-to-face or online; content of the other items on the questionnaire; level of anonymity). These studies should provide increased understanding of the sources of measurement error associated with the Self-Concept Scale along with the factors associated with these errors. This understanding is critical for researchers interested in building and testing theories related to self-concept and for clinicians using these assessments as one source of data for clinical decision making.

References

Anastasi, A. (1982). *Psychological testing* (5th ed.). New York: Macmillan.

Anderson, L. W. (1981). *Affective characteristics in the schools.* Boston: Allyn & Bacon.

Benson, J., & Hocevar, D. (1985). The impact of item phrasing on the validity of attitude scales for elementary school children. *Journal of Educational Measurement, 22,* 213-240.

Bolin, B. L., & Dodder, R. A., (1990). The affect balance scale in American college population. *The Journal of Social Psychology*, 130(6), 839-840.

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81-105.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment.* Beverly Hills, CA: Sage.

Chang, G., Chen, Y.-H., Chen, C.-Y., & Jang, M.-C. (1997). *Relationship of Teacher Criticism to Students' self-concept in Elementary School*. Research Report (NSC-87-2413-H-133-008). Taepei: National Science Council.

Cronbach, L. J. (1950). Further evidence on response sets and test design. *Educational and Psychological Measurement, 10,* 3–31.

DeVellis, R. F. (1991). *Scale development: Theory and applications.* Newbury Park, CA: Sage.

DiStefano, C.,& Motl, R. W. (2006). Further Investigating Method Effects Associated With Negatively Worded Items on Self-Report Surveys. *Structural Equation Modeling: A Multidisciplinary Journal 13,* 440-464.

Hensley, W. E., & Roberts, M. K. (1976). Dimensionality of the Rosenberg self-esteem scale. *Psychological Reports, 78,* 1071-1074.

Horan, P. M., DiStefano, C., & Motl, R. W. (2003). Wording Effects in Self-Esteem Scales:

Methodological Artifact or Response Style? *Structural Equation Modeling: A Multidisciplinary Journal, 10,* 435-455.

Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55.

Kaufman, P., Rasinski, K. A., Lee, R., & West, J. (1991). *National Education Longitudinal Study of 1988. Quality of the responses of eighth-grade students in NELS88.* Washington, DC: U.S. Department of Education.

Kelloway, E. K., Catano, V. M., & Southwell, R. R. (1992). The construct validity of union commitment: Development and dimensionality of a shorter scale. *Journal of Occupational and Organizational Psychology* 65, 197-211.

Marsh, H. W. (1986). The bias of negatively worded items in rating scales for young children: A cognitive-developmental phenomenon. *Developmental Psychology, 22*, 37-49.

Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifacts? *Journal of Personality and Social Psychology, 70,* 810-819.

Marsh, H. W., & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and application* (pp. 177-198). Thousand Oaks, CA: Sage.

Mehrens, W., & Lehmann, I. (1983). *Measurement and Evaluation in Education and Psychology* (3rd ed.). New York: Holt, Rinehart & Winston.

Motl, R.W., & DiStefano, C. (2002). Longitudinal invariance of self-esteem and method effects associated with negatively worded items. *Structural Equation Modeling, 9,* 562–578.

Muthén, L. K., & Muthén, B. O. (1998-2004). *Mplus User's guide*. Third Edition. Los Angeles, CA: Muthén & Muthén.

Nunnally, J. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, M. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88,* 879-903.

Quilty, L. C., Oakman, J. M., & Risko, E. (2006). Correlates of the Rosenberg Self-Esteem Scale method effects. *Structural Equation Modeling, 13,* 99-117.

Roberts, R. E., Lewinsohn, P. M., & Seeley, J. R. (1993). A brief measure of loneliness suitable for use with adolescents. *Psychological Reports*, 72, 1379-1391.

Rosenberg, M. (1965). *Society and the adolescent self-image.* Princeton, NJ: Princeton University Press.

Schmitt, D. P., & Allik, J. (2005). Simultaneous administration of the Rosenberg self-esteem in 53 nations: Exploring the universal and culture-specific features of global self-esteem. *Journal of Personality and Social Psychology, 89*, 623-642.

Schriesheim, C.A., Eisenbach, R.J. & Hill, K.D. (1991). The effect of negation and polar opposite item reversals on questionnaire reliability and validity: An experimental investigation. Educafional *and Psychological Measurement, 51: 67-78.*

Spector, P. E. (1992). *Summated rating scale construction: An introduction.* Newbury Park, CA: Sage.

Tomás, J. M., & Oliver, A. (1999). Rosenberg's self-esteem scale: Two factors or method effects.

Table 1
Chinese General Self-Concept Scale and Descriptive Statistics for Total Sample and by Grade

	Total (n = 752)				Grade 3 (n = 324)				Grade 6 (n = 428)			
	M	*SD*	Skew.	Kurtosis	*M*	*SD*	Skew.	Kurtosis	*M*	*SD*	Skew.	Kurtosis
1. Generally speaking, I like myself.	3.24	0.76	-0.68	-0.15	3.44	0.73	-1.15	0.66	3.08	0.74	-0.44	-0.16
2. I feel I am a happy person.	3.22	0.83	-0.81	-0.12	3.34	0.82	-1.10	0.50	3.14	0.82	-0.62	-0.36
3. I am satisfied with myself.	2.97	0.86	-0.40	-0.64	3.10	0.91	-0.65	-0.54	2.88	0.80	-0.24	-0.54
4. I believe I will be very successful.	2.95	0.91	-0.40	-0.79	3.17	0.89	-0.82	-0.23	2.78	0.88	-0.14	-0.81
5. I do not do most things well.*	2.76	0.80	-0.34	-0.25	2.72	0.89	-0.27	-0.64	2.79	0.72	-0.36	0.10
6. I feel I am not a promising person.*	3.30	0.81	-1.06	0.59	3.19	0.93	-0.93	-0.11	3.39	0.70	-0.98	0.79
7. I feel my academic performances are not as good as others.*	2.72	0.87	-0.26	-0.59	2.70	0.89	-0.15	-0.73	2.74	0.86	-0.34	-0.46
8. My behavior always doesn't meet others' expectation.*	2.87	0.88	-0.39	-0.57	2.84	0.96	-0.33	-0.90	2.89	0.82	-0.43	-0.24
Total	3.00	0.43	-0.17	0.48	3.06	0.44	-0.45	1.27	2.96	0.43	0.03	0.10
Positive	3.10	0.60	-0.39	-0.33	3.26	0.58	-0.78	0.46	2.97	0.58	-0.17	-0.44
Negative	2.91	0.55	-0.33	0.30	2.86	0.57	-0.13	-0.08	2.95	0.53	-0.49	0.76
Cronbach Alpha Positive	.67				.63				.67			
Cronbach Alpha Negative	.54				.47				.61			
Cronbach Alpha Total	.61				.56				.65			

Note. Response scale ranged from 1 (strongly disagree) to 4 (strongly agree).

* Negatively worded items. Response scales were reversed before computing descriptive statistics.

Table 2
Factor Pattern Coefficients for Two-Factor Principal Axis Exploratory Factor Analysis with Promax Rotation (n = 752)

	Factor 1	Factor 2
Positive Item 1 (P1)	.736	-.070
Positive Item 2 (P2)	.702	-.029
Positive Item 3 (P3)	.655	.116
Positive Item 4 (P4)	.752	-.026
Negative Item 5 (N5)	.017	.611
Negative Item 6 (N6)	.190	.609
Negative Item 7 (N7)	-.158	.660
Negative Item 8 (N8)	-.020	.737

Table 3

Goodness-of-Fit Indices for Confirmatory Factor Analysis Models of a Chinese General Self-Concept scale

Model		X^2	df	CFI	RMSEA	SRMR
One-Factor (Model 1a)						
	Grade 3	79.298	20	0.726	0.096	0.078
	Grade 6	155.972	20	0.691	0.126	0.096
	Total	221.067	20	0.699	0.116	0.090
Two-Factor (Positive + Negative) (Model 1b)						
	Grade 3	31.972	19	0.94	0.046	0.047
	Grade 6	38.747	19	0.955	0.049	0.045
	Total	51.313	19	0.962	0.048	0.037
Two-Factor (General + Negative) (Model 1c)						
	Grade 3	22.87	16	0.968	0.036	0.032
	Grade 6	21.234	16	0.988	0.028	0.027
	Total	39.224	16	0.965	0.044	0.028
Two-Factor (General + Positive) (Model 1d)						
	Grade 3	31.807	16	0.927	0.055	0.046
	Grade 6	36.072	16	0.954	0.054	0.043
	Total	48.442	16	0.951	0.052	0.035
One-Factor (General + Correlated Error Negative) (Model 1e)						
	Grade 3	21.962	14	0.963	0.042	0.031
	Grade 6	16.159	14	0.995	0.019	0.023
	Total	34.769	14	0.969	0.044	0.026
One-Factor (General + Correlated Error Positive) (Model 1f)						
	Grade 3	27.583	14	0.937	0.055	0.044
	Grade 6	33.224	14	0.956	0.057	0.042
	Total	41.163	14	0.959	0.051	0.033

Note: Total *N*=752, Grade 3 *n*=324, Grade 6 *n*=428

CFI = Comparative fit index; SRMR = Standardized root mean square residual; RMSEA = Root mean square error of approximation.

Table 4
Standardized Parameter Estimates for Confirmatory Factor Analysis Models (*n* =752)

	Model 1a One-Factor	Model 1b Two-Factor		Model 1c Two-Factor		Model 1d Two-Factor	
	General Self-Concept	Positive Self-Concept	Negative Self-Concept	General Self-Concept	Negative Method	General Self-Concept	Positive Method
P1	0.573	.589		.581		.140	.575
P2	0.511	.519		.518		.133	.500
P3	0.594	.581		.591		.215	.536
P4	0.629	.651		.648		.150	.640
N5	0.151		.389	.098	.368	.397	
N6	0.287		.530	.238	.455	.546	
N7	0.101		.407	.038	.431	.401	
N8	0.189		.596	.121	.612	.578	
		r = .264		r = 0[a]		r = 0[a]	

[a] Correlation fixed to zero.

Table 4 (continued)
Standardized Parameter Estimates for Confirmatory Factor Analysis Models (*n* =752)

	Model 1e One-Factor					Model 1f One-Factor				
	General Self-Concept	Correlated Uniqueness Negative				General Self-Concept	Correlated Uniqueness Positive			
P1	.581					.141		P1	P2	P3
P2	.519					.129	P2	.251		
P3	.589					.214	P3	.306	.312	
P4	.649					.152	P4	.391	.318	.319
N5	.093		N1	N2	N3	.398				
N6	.236	N2	.211			.546				
N7	.041	N3	.148	.169		.401				
N8	.122	N4	.205	.274	.285	.577				

Table 5

Fit Indices from Factorial Invariance Tests of One-Factor Model (Model 1e) with Correlated Uniquenesses for Negatively Worded Items for Grades 3 and 6

Model	χ^2	df	$\Delta\chi^2$	Δ df	CFI	RMSEA	SRMR
1. No Equality Constraints	38.122	28	--	--	.985	.031	.027
2. Equal Factor Pattern Coefficients	44.515	35	6.393	7	.986	.027	.033
3. Equal Covariances Between Uniquenesses	49.296	41	4.781	6	.987	.023	.036

CFI = Comparative fit index; SRMR = Standardized root mean square residual; RMSEA = Root mean square error of approximation.

None of the $\Delta\chi^2$ was statistically significant at the .05 level.

Table 6

Fit Indices from Factorial Invariance Tests of Two-Factor Model (Model 1c) with One General Self-Concept Factor and Negative Method Factor for Negatively Worded Items for Grades 3 and 6

Model	χ^2	df	$\Delta\chi^2$	Δ df	CFI	RMSEA	SRMR
1. No Equality Constraints	44.104	32	--	--	.982	.032	.029
2. Equal Factor Pattern Coefficients for General Self-Concept Factor	50.012	39	5.908	7	.983	.027	.035
3. Equal Factor Pattern Coefficients for Negative method Factor	54.874	42	4.862	3	.980	.029	.038

CFI = Comparative fit index; SRMR = Standardized root mean square residual; RMSEA = Root mean square error of approximation.

None of the $\Delta\chi^2$ was statistically significant at the .05 level.

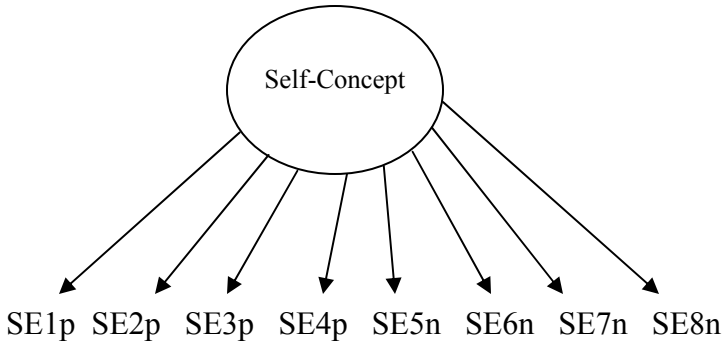

FIGURE 1a One-factor self-esteem.

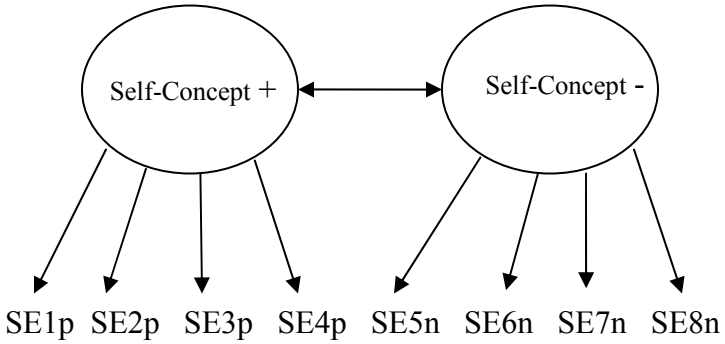

FIGURE 1b Two-factor self-esteem (positive and negative substantive factors).

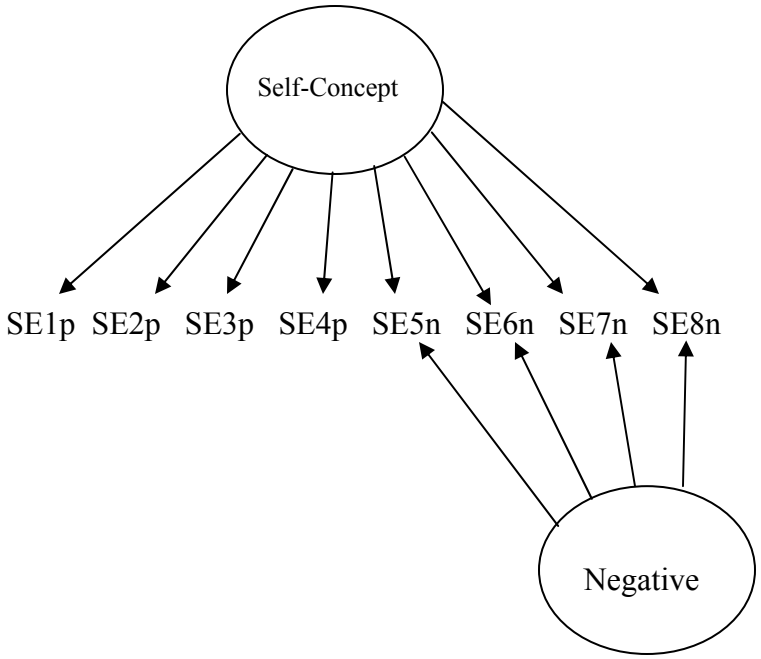

FIGURE 1c Two-factor self-esteem (general and negative wording factor).

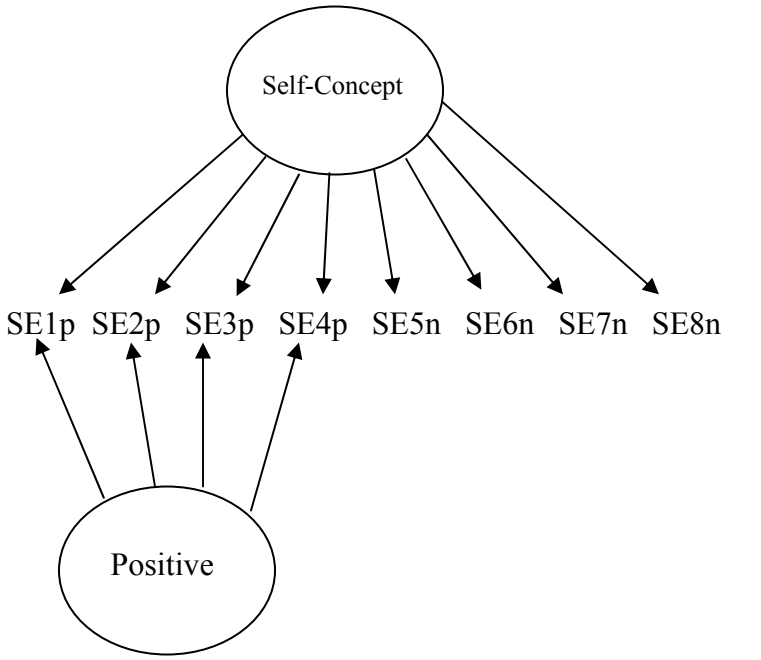

FIGURE 1d Two-factor self-esteem (general and positive wording factor).

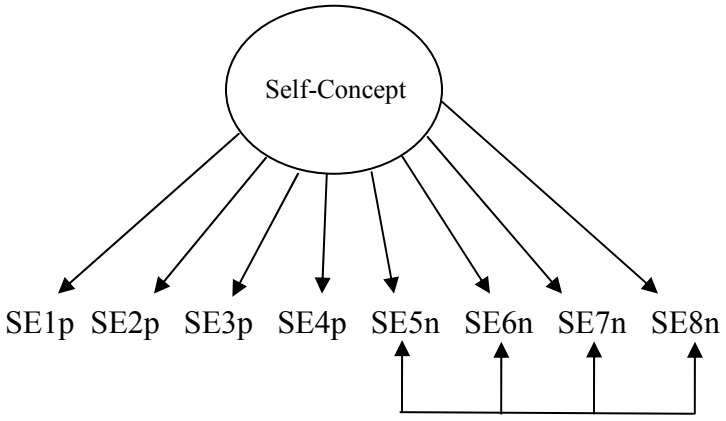

FIGURE 1e One-factor self-esteem with correlated error for negatively worded items.

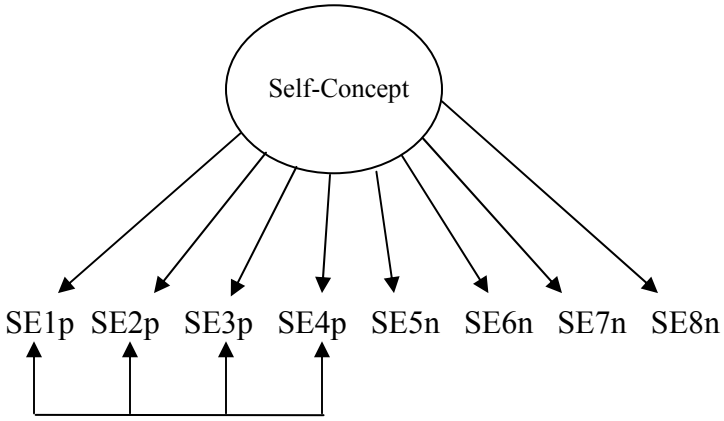

FIGURE 1f One-factor self-esteem with correlated errors for positively worded items.

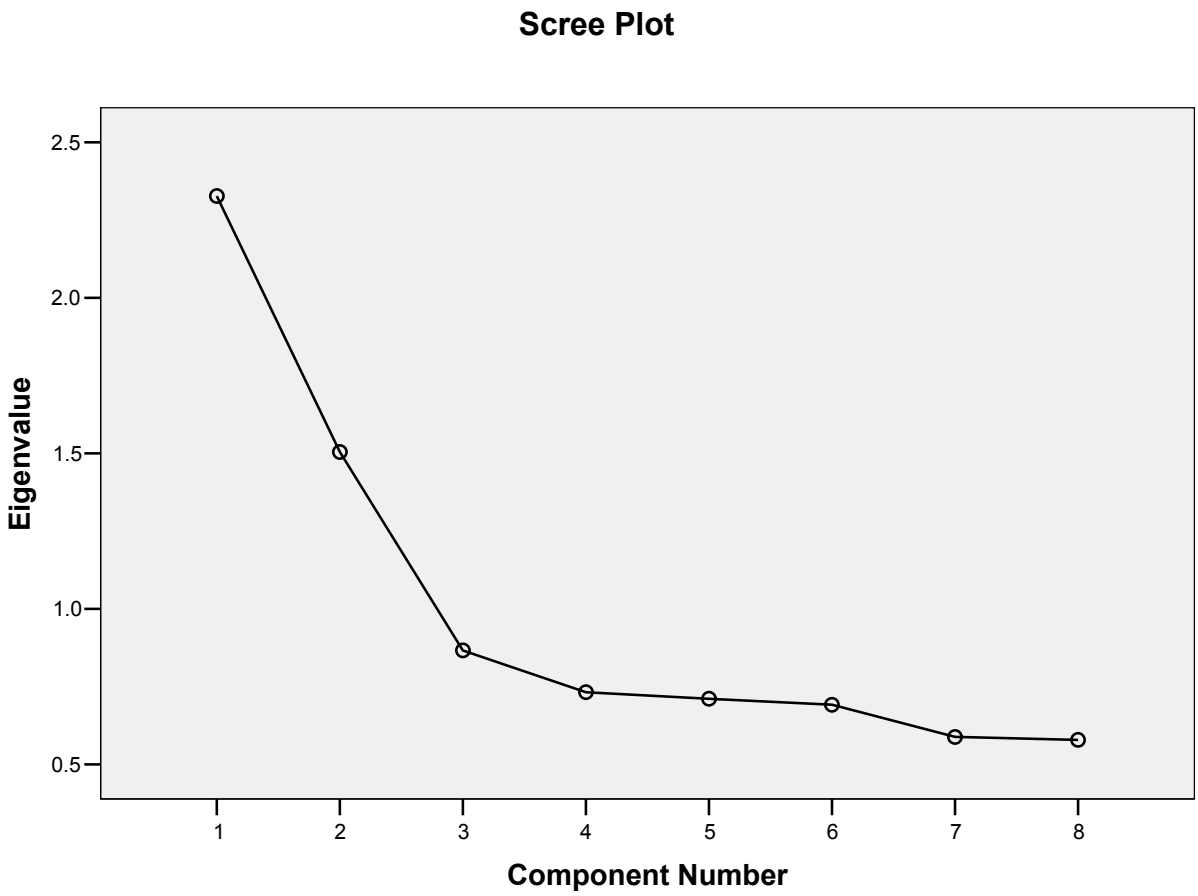

Figure 2 Scree plot for the Chinese general self-concept scale

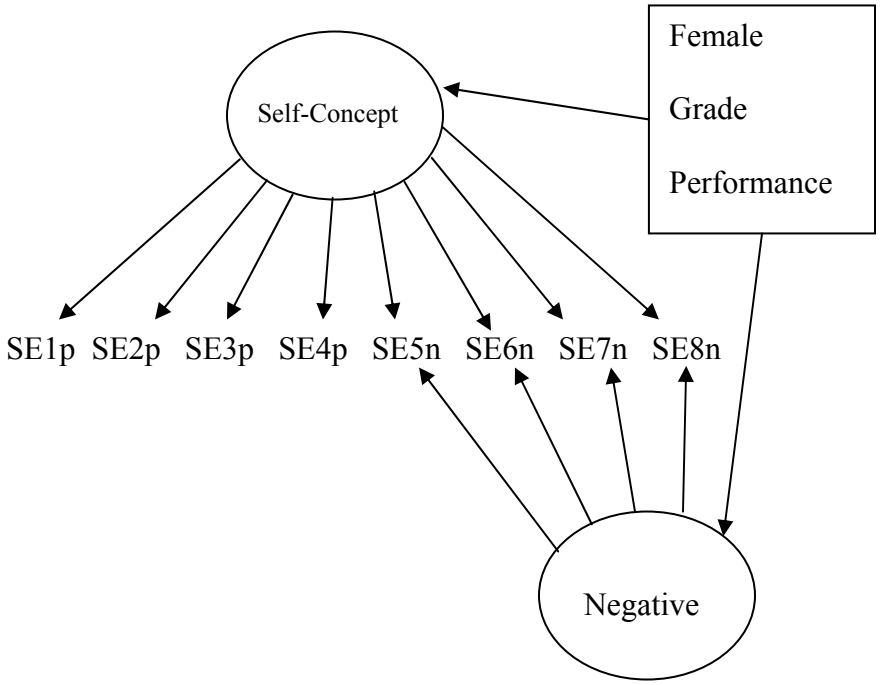

Figure 3 Path model with gender (female =1, male = 0), grade level (3, 6), and teachers' ratings of students' performance (1=high, 2=average, 3=low) predicting general self-concept and negative method effects.